Information Fusion 24 (2015) 137–146

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Cokriging for cross-attribute fusion in sensor networks

Javier Palafox-Albarran [a,b,*], Reiner Jedermann [a,c], Bonghee Hong [d], Walter Lang [a,b,c]

[a] Institute for Microsensors, -Actuators and -Systems (IMSAS), University of Bremen, Otto-Hahn-Allee NW1, Bremen, Germany
[b] Bremen Research Cluster for Dynamics in Logistics (LogDynamics), Bremen, Germany
[c] Microsystems Center Bremen (MCB), Bremen, Germany
[d] Dept. of Computer Engineering, Pusan National University, Pusan National University, Busan 609-735, Republic of Korea

ARTICLE INFO

Article history:
Received 26 February 2013
Received in revised form 18 July 2014
Accepted 19 September 2014
Available online 30 September 2014

Keywords:
Wireless sensor networks
Cokriging
Coregionalisation model
Cross-attribute fusion
Submodularity

ABSTRACT

Optimisation of the number of required measurement points and their location is an important research topic in sensor networks. Finding the optimal positions increases spatial coverage and reduces deployment costs. This paper presents an approach for the case that two attributes have to be measured with a different number of available sensors. The proposed cokriging method performs cross-attribute fusion in sensor networks by being based on the analysis of multi-variable spatial correlations. To the best of our knowledge, this scientific work is the first one considering kriging and cokriging interpolations as IF methods. The single-variable ordinary kriging and bi-variable methods were applied to experimental data. The combination of humidity and temperature data in a refrigerated container is used as exemplary case, humidity measurements are considered to be the expensive attribute to measure. The average estimation error for intermediate points was estimated as a function of the number of humidity sensors. When variability is high, data fusion using the bi-variable method produced results as accurate as the single-variable one, without the necessity of deploying a large number of humidity measuring points, by complementing the estimation with temperature measurements.

1. Introduction

"Wireless sensor networks" is a promising technology that consists of small, cheap devices that have a powerful combination of sensing, computing and communication capabilities. They are deployed to monitor specific attributes of interest associated with measurements of different physical properties. Ideally, to maintain profitability, they must be able to cover an area of interest using the minimum possible number of sensors. A dense deployment of all required sensors types is not feasible; in some cases, the sensor type required to monitor one of the attributes might be too expensive to deploy even in low numbers. One possible way to improve the coverage without increasing deployment costs is by finding the optimal positions of the sensing points and by repositioning them. Unfortunately, in some applications such as cold-chain monitoring, repositioning of the sensors is not possible after deployment.

To solve this problem, we propose the use of Kriging methods, which are Geostatistical techniques, on the wireless sensor networks (WSNs) field. The Kriging methods can be regarded as Information Fusion (IF) methods, but they have not been fully appreciated by the general fusion community. The methods are based on statistical descriptions of the spatial dependencies of the attributes, the so-called variograms. They are Best Linear Unbiased Estimators (BLUE), and fuse the available data, and the relationship between the different sources to provide not only and estimate of the attributes at a specific location but also its uncertainty, the so-called kriging variance (KV).

The Ordinary Kriging method (OK) which considers a single attribute can be considered as a statistical fusion method across sensors. Cokriging interpolation (CK) can be considered a method for fusion across attributes and sensors. In the context of sensor networks it might bring some interesting advantages: according to [1], adding data from a secondary attribute to estimate the primary one, could only increase precision; the so-called cokriging variance can only be less than or equal to the KV of the same primary attribute; which also means that the certainty of the fusion is

* Corresponding author at: Institute for Microsensors, -Actuators and -Systems (IMSAS), University of Bremen, Otto-Hahn-Allee NW1, Bremen, Germany. Tel.: +49 (0)421 218 62637; fax: +49 (0)421 218 98 62637.

E-mail addresses: jpalafox@imsas.uni-bremen.de (J. Palafox-Albarran), rjedermann@imsas.uni-bremen.de (R. Jedermann), bhhong@pusan.ac.kr (B. Hong), wlang@imsas.uni-bremen.de (W. Lang).

http://dx.doi.org/10.1016/j.inffus.2014.09.007
1566-2535/

improved. More importantly, when one of the attributes which is not sampled at every location, as might be the case of expensive sensors, the use of a secondary attribute which is sampled densely, allows an estimate of a primary under-sampled attribute at the non-sampled locations without necessity of repositioning the deployed sensors For instance, instead of deploying 50 humidity sensors, collocated temperature sensors might help estimate humidity levels for 40 of the positions by using only 10 genuinely acquired measurements.

The goal of this research is to study the advantages of CK interpolation as an IF technique in the context of WSNs. Our exemplary case is the environmental attributes inside a refrigerated container where mobility of the nodes is not possible; humidity is considered as the expensive attribute that must be estimated with few sensors whereas temperature is densely sampled. Due to the strong theoretical guarantees through a calculation of the variance of the estimates, the applicability of results presented in this paper can be generalised to other applications and scenarios.

Our first aim is to verify whether it brings any significant advantage in the estimation accuracy when compared to OK. The second aim is to determine the minimum number of temperature and humidity sensors required in order to accurately interpolate a spatial temperature profile when humidity is chosen as the primary attribute. This paper begins by giving a review of the related work and presenting the experimental data. Section 4 mentions the basic steps of the cokriging interpolation method. The following section explains the notion of cross-attribute submodularity, an interesting property of cokriging that allows increasing spatial coverage using any kind of correlated information. Section 6 presents the main simulation results for the exemplary case of a refrigerated container. In Section 7, the suitability of their deployment in a WSN gateway are tested. Java based OSGi (formerly Open Source Gateway Initiative) implementations of the required algorithms are tested and compared. Finally conclusions are made.

2. Related work

In recent years, much research has been made on the development of intelligent containers for cargo security and integrity. Most of the research focuses on information and communication technologies (ICT) for tracking, tracing and monitoring of the goods. Sensor networks research on containers can be found for example in the research performed by Jedermann et al. [2], he proposed to equip the containers with RFID data loggers with sensing and processing capabilities to estimate shelf-life. In the research work of Carullo et al. [3], sensor networks were used to monitor environmental quantities such as temperature, humidity, and air flow to detect local deviations that might be an indicator of an unauthorized opening of the doors or inadequate cooling. Regarding commercial projects, an example is the project by IBM called Secure Trade Lane [4] which provides a solution for more predictable container shipments.

Jedermann et al. [5,6] recognised the problem of lack of coverage when measuring temperature inside a refrigerated container and proposed the use of Kriging interpolation, which is commonly used in geological sciences, to estimate the temperature variations in refrigerated trucks in the context of sensor networks but he limits his research to the interpolation of one single attribute. Another example of considering Kriging interpolation in WSN is the work of Umer et al. [7], who focuses on developing distributed algorithm for Kriging interpolation in resource constrained sensor nodes but nothing is mentioned about robustness of the estimates or their uncertainty.

Kriging interpolation for sensor repositioning in WSN can be found in [8,9]. They only focus on single attribute interpolation, even though their studies required estimation of more than one spatial phenomenon. The authors in [8] follow an information driven approach for sensing optimisation; they find the optimal positions of the sensor to extract the maximum information. They use Kriging interpolation to determine the best position of the new measurement locations, the sensors are allowed to move until convergence is achieved. Krause and Guestrin's et al. mobility approach [9] propose to minimise the Kriging variance in the unobserved locations using only the minimum required observations. They mentioned important properties of kriging interpolation: *strong theoretical guarantees and submodularity*. The first one ensures the general validity of the results through the Kriging variance (KV), whereas, the second one allows to find an optimal number for the most expensive sensors.

More related works can be found in geostatistics literature. For example, Szidarovszky et al., in [10] proposed an off-line Branch-and-Bound algorithm for the estimation of the minimum number of sensors. The method minimises the number of required additional points subject to upper bounds of the Kriging-Variance. Several research papers combing IF and Geographical information systems (GIS) can be found in [11]; surprisingly, kriging interpolation, developed for geostatistics, is not considered.

3. Experimental data

The interpolation procedures were tested on two datasets recorded in a refrigerated container of dimensions $2.2 \times 2.2 \times 5.4$ metres within a collaborated internship with the Research Centre for Logistics Information Technology (LIT), at the Pusan National University in Korea, in 2012.

In order to increase spatial variability, the container was first cooled from ambient temperature (15 °C) to a set point of 0 °C for 3 h, then warmed to a set point of 25 °C for about 2 h, prior to performing the actual experiment, which comprised of cooling the container to 5 °C for two hours and warming it to a set point of 20 °C for another two hours. During the 4 h that the experiment lasted, N_k samples were acquired for both attributes at regular intervals. As the most significant influence on the cargo container is the loading state, two configurations were tested: one with an empty container and the other with pallets covering the floor to deflect the air flow. Fig. 1 shows the average temperature and humidity measurements and the positions of the sensors for the experiment with pallets on the floor.

In total, 60 ASN 405T [12] wireless sensor nodes were placed in the walls, doors, ceiling and floor forming a grid of 55 cm lag distance. Each node contains a SHT20 [13] sensor capable of measuring humidity and temperature, sent to a gateway with an accuracy of ±3% RH and ±3 °C for humidity and temperature, respectively.

The measurement data set was split into two groups. The first group of measurements serves as the input for data fusion. The second group serves as a reference; these measurements at the destination points z_i were compared to the estimates $\hat{z}_i$. In order to be consistent with our previous research and with the aim of comparability, the error ε_i for one destination point i is defined as the root-mean square deviation between the estimated $\hat{z}_i(k)$ and the measured value $z_i(k)$ over all samples N_k for the test duration.

$$\varepsilon_i = \frac{\sum_{k=1}^{N_k} (\hat{z}_i(k) - z_i(k))^2}{N_k} \tag{1}$$

$\bar{\varepsilon}$ is the average of all errors ε_i.